An Ordinal Scaling Method for Questionnaire and Other Ordinal Data

Norman Cliff and Linda M. Collins
University of Southern California

Judith Zatkin
Southern California Edison

Dannie Gallipeau
University of Southern California

Douglas J. McCormick
American Telephone and Telegraph

This paper reports the development and application of a method for ordering persons and items (or stimuli) when responses are ordinal. The method applies most directly to data where responses are dichotomous, indicating agreement or acceptableness or similarity, and can be assumed to reflect proximity rather than dominance. It orders rows and columns of the response matrix into "parallelogram" form, using pairwise interchange procedures, followed by other steps. The method was applied to several sets of questionnaire data and one set of archeological data, with reasonable success. Other applications and extensions are suggested. *Index terms: Dichotomous responses, Interchange methods, Ordinal scaling, Parallelogram scaling, Proximity data, Questionnaire responses.*

The present research was devoted to the development and implementation of methods for analyzing proximity scales. These methods are often called "proximity scaling" (Coombs, 1964) and also often misnamed (Carroll & Arabie, 1980; Heiser, 1981) "unfolding" analysis. In proximity scaling the general idea is that there is an underlying continuum, and a *1* indicates that two things are close on the continuum and a *0* indicates that they are distant.

With the typical person × stimulus matrix it is assumed that there is for each person an ideal point on the stimulus or item continuum, and she/he accepts or endorses those items that are near the ideal point (*1*) and does not endorse those that are distant (*0*). If this is a correct assumption, and if there is a complete matrix of such binary data indicating proximity/distance between the row entity and the column entity, it should be possible to order the rows and the columns so that the *1*s form a diagonal stripe from one corner of the matrix to the opposite corner. These scales are often called "parallelogram scales" because of the roughly parallelogram shape of this stripe.

The "unfolding" label arises from the fact that analysis of pairwise dominance data for the stimuli (columns) will result in a folded-over version of the continuum, which then needs to be "unfolded." However, pairwise dominance is not really relevant to parallelogram scales. Both Guttman scale and parallelogram scale items can be ordered along a continuum representing the underlying construct. In the case of Guttman scales, item dominance provides this order, while item dominance has no implications for this order in parallelogram scales. Nonetheless, the term "unfolding" has come to be attached to this type of data (Carroll & Arabie, 1980; Heiser, 1981).

Ordinal Datasets

One of the goals of the present study was to explore the hypothesis that persons respond to most questionnaire items in terms of the proximity of an item's meaning to some "ideal point." This ideal point may represent a person's own point of view, or the point of view a person wishes to present if dissimulation (intentional or otherwise) is occurring. When the meaning of the item is "close" to a person's ideal point, there is a positive response; otherwise a negative response is given.

Most attitude and temperament scales are constructed so as to "measure" the presumed underlying trait or attitude continuum by counting the number of items that are responded to in the "keyed" direction. For example, on an attitude-toward-war scale, there would be statements favorable and unfavorable toward war. Statements favorable to war are keyed so that positive responses (i.e., endorsements) are keyed positively, while those that are against war are keyed negatively (i.e., disagreement or non-endorsement is keyed positively). Thus high scores are obtained by persons who endorse statements favorable to war and reject statements unfavorable to it. Low scores are obtained by persons who do the reverse, and middle scores are obtained by those who do some of each.

Items typically are selected for such scales by means of a correlational strategy, with the items (whether positively or negatively keyed) that correlate positively with each other included in the scale. Items with low correlations are discarded. When the proximity model holds, this means that items toward the middle of the set tend to be discarded because they are curvilinearly related to the underlying continuum, and endorsement of them correlates neither positively nor negatively with the other items. Thus it is possible that commonly-used scale construction methods mask a simple response model that could be used with such data.

Two-set proximity data can be expected to arise in a number of contexts other than attitude or self-description data. For example, the voting records of members of a legislative body frequently follow this form. The data consist of a binary matrix of legislators $\times$ bills, and a *1* means that the legislator voted for the bill. It is often observed that such data approximate the parallelogram form, with the underlying continuum presumed to represent a liberal-conservative dimension. Also, archeological data (where the matrix is sites $\times$ artifact types) are likely to show this form, with *1* meaning the type was found at the site, and the underlying continuum representing time.

Typically, this type of data has been analyzed by proposing a quantitative model relating the probability of a *1* to the distance between a scale value for the row entity and the scale value of the column entity. Alternatively, it has been analyzed by a rather ad hoc application of, for example, multidimensional scaling to the data matrix X or to $X'X$ and/or some related matrix (see Kendall, 1971). This may be successful in the sense of providing heuristically useful information, but it seems desirable to seek methods that do not impose a rather arbitrary quantitative structure on the data, but rather retain its simply ordinal nature. The data seem to represent merely ordinal relations; those items for which the response is *1* are simply closer to the person's ideal point than those marked *0*. Thus it would be preferable to apply only ordinal methods to the analysis of these data. The development of such ordinal methods was the major objective of the present research.

Parallelogram Scaling

Criterion Functions

In the ideal case of parallelogram scaling, it is possible to reorder the rows and columns of a binary data matrix so that the parallelogram form is shown. For example, given the matrix in the left panel of Table 1, it is possible to permute rows and columns to achieve the form in the right panel of the table.

Table 1
Scrambled and Reordered Ideal
Parallelogram Scale

	Scrambled		Reordered	
	Items		Items	
	0 1 1 0 0 0 0 0 0 1		1 1 1 0 0 0 0 0 0 0	
	0 1 1 0 0 1 0 0 0 0		0 1 1 1 0 0 0 0 0 0	
	0 0 1 0 0 1 0 0 1 0		0 0 1 1 1 0 0 0 0 0	
Persons	0 0 0 1 0 1 0 0 1 0	Persons	0 0 0 1 1 1 0 0 0 0	
	1 0 0 1 0 0 0 0 1 0		0 0 0 0 1 1 1 0 0 0	
	1 0 0 1 0 0 1 0 0 0		0 0 0 0 0 1 1 1 0 0	
	1 0 0 0 1 0 1 0 0 0		0 0 0 0 0 0 1 1 1 0	
	0 0 0 0 1 0 1 1 0 0		0 0 0 0 0 0 0 1 1 1	
	0 0 0 0 1 0 0 1 0 0		0 0 0 0 0 0 0 0 1 1	
	0 0 0 0 0 0 0 1 0 0		0 0 0 0 0 0 0 0 0 1	

The problem is to find the permutations that achieve this result. The difficulties are exacerbated in the case of real data because the presence of error or inconsistency lowers the likelihood of achieving the exact parallelogram form.

Modern psychometric methods revolve around selection of some numerical characteristic of the data that measures its departure from the ideal, and attempting to find parameter values, in this case the orders, that optimize this objective function (see, e.g., Cliff, 1973; Guttman, 1971). The problem in the present case is then one of selecting the objective function and finding a means of optimizing it.

The most obvious characteristic that distinguishes the left and right panels of Table 1 is the fact that within any row (or column, but the present discussion will refer only to rows), the 1s are bunched together in the right panel and scattered in the left one. Therefore, some measure of the average scatter within rows seems appropriate as an objective function. One appropriate measure of spread is the variance of the ranks of the 1s in a row. That is, let x_{ij} be the binary response in row i column j of the matrix. Then the variance of the ranks of the 1s in a row is

$$s_i^2 = \frac{\sum_j (jx_{ij} - \bar{j}_i)^2 x_{ij}}{x_{i.}} \quad , \tag{1}$$

where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, p$,

$$x_{i.} = \sum_j x_{ij} \quad , \tag{2}$$

and

$$\bar{j}_i = \frac{\sum_j jx_{ij}}{x_{i.}} \quad . \tag{3}$$

The extra x_{ij} term on the right side of Equation 1 is simply for the purpose of including in the sum only those differences where $x_{ij} = 1$. The goal is to find the permutation of the columns such that the average of these row variances is a minimum:

$$\phi = \frac{s_i^2}{n} \quad . \tag{4}$$

Rather than actually permuting the columns, it is simpler to permute the column indices. Let these be

denoted r_j. Also, for the present purposes it is helpful to separate Equation 1 into the difference between a sum of squares and a squared sum. This allows, among other things, the omission of the extra x_{ij}; Equation 1 then becomes

$$s_i^2 = \frac{\sum r_j^2 x_{ij}}{x_{i.}} - \frac{\left(\sum r_j x_{ij}\right)^2}{x_{i.}^2} \quad . \tag{5}$$

In Equation 5, then, r_j is the ordinal position or rank of column j in the current permutation of the column indices.

Some other measures of spread were considered. For example, another indication of a poor order is that the *1*s and *0*s are interspersed (e.g., 01011001 as opposed to 01111000). Therefore, the number of runs of symbols of one kind might be used as a criterion. This has a disadvantage in the error-perturbed case, namely that it does not distinguish between 10011001 and 01011010, whereas the latter is clearly preferable. Similarly, a criterion such as minimizing the number of *0*s that are surrounded by *1*s has the disadvantage that it is solely dependent on the positions of the two outer *1*s. A number of possibilities of this type were considered but discarded on the basis of such considerations.

One drawback to a least-squares criterion such as Equation 4 is that squaring imposes a very large penalty for large deviations. A criterion such as minimizing the sum of absolute deviations from the median might therefore be preferable. Minimizing such a function is very complex, and was not attempted in the present case.

Minimizing ϕ

There are a variety of means of attempting to minimize the objective function ϕ. However, note that the solution for the r_j is limited here to the case where they are consecutive integers. This means that this is a form of the "traveling salesman" problem (Baker & Hubert, 1977; Hubert, 1974) and therefore there is no guarantee of finding the minimum short of trying all $p!$ permutations. However, heuristically based algorithms may be successful in most cases.

The simplest such algorithm is pairwise interchange, and this was the primary approach here. The effect of interchanging every pair of columns was computed, the interchange that gave the largest reduction in ϕ was selected, and that pair was interchanged. Then the effects of a second interchange were computed, the most effective next interchange was found, and the process was continued until no interchange would yield an improvement. The resulting order was taken as the optimum one in the initial version of the program.

Computing the effect on ϕ of interchanging all $\frac{1}{2}p(p-1)$ pairs of columns seems like a formidable task, and repeating the process after each interchange compounds the complexity. However, examination of the functions involved in Equation 4 shows that the major part of the computation was rather simple and only had to be performed once. Furthermore, updating the effects of potential interchanges is simplified when it is realized that this need only be done for pairs involving the columns that were interchanged in the preceding step. The details of these computations are given below.

This procedure was tried out on a number of simple errorless examples where the correct solution was known, and it achieved the minimum in a few steps each time. However, when it was used on real data where there was no perfect solution, some applications worked well, but in others the results differed from what would be expected on substantive grounds. These tended to be of a particular form. In these unsatisfactory cases, the items were mostly in the order that would be expected on substantive grounds, but it was as if some items at one end of the scale had been attached to the wrong end. For example, a sequence that was expected to be 123456789 ended up as 456789123.

Consequently, an additional step was added to the procedure; it takes place after no further inter-changes improve the criterion. This step is a type of permutation in which the k items at one end of the scale are moved to the other end without changing their order, and it is thus termed "break-and-rejoin". The operation can be expressed as

$$r'_j = r_j + k \ (\text{modulo } p) \quad . \tag{6}$$

This is done for $k = 1, 2, ..., p - 1$. If any such permutation results in an improvement in ϕ, then it is used, and the pairwise interchange procedure is again applied to the resulting order. This process continues until neither procedure results in an improvement.

Despite the efficiencies that were introduced into the pairwise interchange process, the amount of computing time still increases rapidly with p, the number of indices to be interchanged. Therefore, a procedure that would yield a starting order was found. (This procedure relates the present problem to a variety of others, as will be discussed below.) These developments resulted in two computer programs written in APL and the PROC MATRIX system of SAS.

Specific Procedures

Operations designed to minimize Equation 4 are clarified if some alterations are made in Equation 5. First, note that for binary data $x^2_{ij} = x_{ij}$, and make this substitution in the sum on the left side of Equation 5. Also,

$$\frac{\left(\sum_j r_j x_{ij}\right)^2}{x^2_{i.}} = \left(\frac{\sum_j r_j x_{ij}}{x_{i.}}\right)^2 \tag{7}$$

and let $y_{ij} = x_{ij}/x_{i.}$. Therefore

$$s^2_i = \sum_j r^2_j y_{ij} - \left(\sum_j r_j y_{ij}\right)^2 \quad . \tag{8}$$

Then

$$\phi = \frac{\sum_i \sum_j r^2_j y_{ij} - \sum_i \sum_j \sum_k r_j r_k y_{ij} y_{ik}}{n} \quad . \tag{9}$$

Changing the order of the summations and then factoring out terms according to their indices leads to

$$\phi = \frac{\sum_j r^2_j \sum_i x_{ij} - \sum_j r_j \sum_k r_k \sum_i y_{ij} y_{ik}}{n} \quad . \tag{10}$$

In matrix form this may be seen as

$$\phi = r'(D_\Sigma - Y'Y)r \quad , \tag{11}$$

where D_Σ is a diagonal matrix whose entries are $\sum_i y_{ij}$,

r is a vector of the first p integers, and

Y is an $n \times p$ matrix of the y_{ij}.

Then the solution to the problem of minimizing the average variance of the ranks of the items endorsed by individuals becomes one of finding the permutation of the integers to place in r that minimizes Equation 11.

Because of the restriction that r must contain the first p integers, there is no guaranteed solution to the problem short of trying all $p!$ permutations. Therefore, heuristic approaches to finding an acceptable

solution are used. As noted above, a two-step algorithm was adopted, with pairwise interchange and break-and-rejoin as the first and second steps.

Literal calculation of Equation 11 for all pairwise interchanges is itself a formidable task if the value of p is greater than about 12. Several modifications were introduced to increase efficiency. An obvious one is to combine $\mathbb{D}_\Sigma - \mathbb{Y}'\mathbb{Y}$ into a single permanent matrix, denoted $\mathbb{G}$. A number of other efficiencies are possible in the interchange process. Let ϕ_{uv} represent the change in ϕ resulting from interchanging columns u and v. Then

$$\phi_{uv} = (r_u - r_v)\left[2\left(\sum r_j g_{uj} - r_j g_{vj}\right) - (r_u - r_v)(g_{uu} + g_{vv} + 2g_{uv})\right] \quad , \tag{12}$$

where r_u and r_v are the current ranks of u and v. This results in a very substantial saving in computation because the only sum of products involves only one index. It can be calculated at the beginning of the iteration for all pairs. Furthermore, once a pair has been selected for interchange, the only components of the computation that change are those which involve the members of that particular pair. The programs that were written to implement the procedure take advantage of some but not all of these efficiencies, because they require a level of detail that is not well-suited to the high level of languages used.

Starting Orders

The process would be expedited if a rational starting order were available that reduced the number of required steps. The form of Equation 11 is very suggestive. It indicates that if $\mathbb{r}$ were not restricted to consist of a permutation of the first p consecutive integers, minimizing ϕ would become an eigenvalue-eigenvector problem. Indeed, $\mathbb{r}$ would be the eigenvector corresponding to the smallest nonzero eigenvalue of $\mathbb{D}_\Sigma - \mathbb{Y}'\mathbb{Y}$.

At least one eigenvalue is 0 because $\sum_k g_{jk} = 0$, since $\sum_i \sum_k y_{ij} y_{ik} = \sum_i y_{ij}$. To see this, note that

$$\sum_i \sum_k y_{ij} y_{ik} = \sum_i y_{ij} \sum_k y_{ik} \tag{13}$$

but

$$y_{ik} = \frac{x_{ik}}{x_{i.}} \quad . \tag{14}$$

Therefore

$$\sum_k y_{ik} = \frac{\sum_k x_{ik}}{x_{i.}} = 1 \tag{15}$$

and

$$\sum_i \sum_k y_{ij} y_{ik} = \sum_i y_{ij} \quad . \tag{16}$$

Thus, if $\mathbb{r}$ were not restricted to be a permutation of p consecutive integers, it would be the eigenvector corresponding to the smallest nonzero eigenvalue of $\mathbb{G}$, say $\mathbb{r}^*$. The eigenvector corresponding to the zero eigenvalue is the constant vector because Equation 16 shows that $\mathbb{G}\mathbb{1} = 0$. Therefore, $\mathbb{1}'\mathbb{r}^* = 0$ because of orthogonality of eigenvectors (i.e., the average element of this $\mathbb{r}^*$ is 0). This suggests that the first p integers, ordered in the same way as $\mathbb{r}^*$, should be an effective starting order for $\mathbb{r}$, and this is incorporated into the procedure.

A second basis for a starting order is the "nearest neighbor" order. The idea underlying this is the fact that in a perfect parallelogram matrix, the pattern of 1s and 0s in any column is "most like" the columns on either side of it. The nearest-neighbor ordering strategy (e.g., Gelfand, 1971) is sometimes

used in the present context to find the approximate parallelogram order. There are, of course, many possible definitions of "most like". Here, the Goodman-Kruskal gamma (Goodman & Kruskal, 1954) is used as the measure of similarity of columns. Gamma is equal to $(ad - bc)/(ad + bc)$, where a is the number of instances where there is a 1 in both columns, d is the number where both are 0, b is the number of $1,0$, and c is the number of $0,1$. Gamma is equal to 0 if the presence of a 1 in one column is independent of the presence of a 1 in the other column; gamma is 1.0 if the 1s in one column are a subset of the 1s in the other (i.e., there may be instances of $1,0$ or $0,1$, but not both).

Nearest-neighbor ordering starts by finding the pair of columns that have the highest gamma. Then a third column is found that is closest to one of these two, and it is attached to the closer column. It is now one end of a three-element chain; then the fourth element is found that is the closest to one of the current ends, and it is added to that end. The process continues, adding elements one at a time to one end or the other until all are ordered. The result is a nearest-neighbor order. This, too, was used as a starting order.

The unrestricted solution to Equation 11 indicates that the present problem is a variation on the problem discussed by Heiser (1981), which is directly related to what has been called the "method of reciprocal averages" (see, e.g., Nishisato, 1980) and also to the extensive French literature on correspondence analysis (Tenenhaus & Young, 1985). Levine (1979) considered a closely related topic, although differently expressed.

These other literatures differ in two ways from the present approach. First, their goal is generally to find scales (r* vectors in the present notation) that maximize between-row variance. Obviously, minimizing within-row variance and maximizing between-row variance are interchangeable goals, hence this is not really a difference. The present procedure, however, minimizes average within-row variances, whereas these other methods are actually maximizing sums of squares. This amounts to a difference in the weighting of the rows of the X matrix, so this is perhaps only a minor difference conceptually. It is one that could lead to substantively important differences in any given application, however. The second and more important difference is that the present method considers only the ordinal characteristics of the data. That is, the solution is restricted to using consecutive integers for r, in keeping with the ordinal character of the observed data.

Illustrative Applications

First Questionnaire Study

This study, which illustrates the process of parallelogram analysis, consisted of the analysis of responses to items chosen to represent several scales. Accordingly, items representing scales of Self-Esteem (SE), Tolerance for Ambiguity (TA), and Attitude Toward War (War) were selected from Robinson and Shauer (1973). These were augmented by additional items written by the authors, particularly where middle-range items were felt to be needed. The result was a set of three scales: SE (13 items), TA (30 items), and War (17 items). Only the results concerning SE and War are discussed here. The SE and War items were intermixed with additional items (which belonged to none of these scales) to form a questionnaire.

The questionnaire was administered to 90 undergraduates fulfilling a course requirement to take part in research studies. The responses were on a four-point scale: Strongly Agree, Tend to Agree, Tend to Disagree, Strongly Disagree. The responses were dichotomized into Agree (1) and Disagree (0) for the first set of analyses.

The items on the Self-Esteem scale are listed in Table 2 in the final order that the program found for them, whereas the item numbers are the initial, arbitrary order in which the items were listed at the

Table 2
Ordered Self-Esteem Items

 4. I feel I can't do anything right.
 5. I feel I do not have much to be proud of.
12. Sometimes I think I'm no good at all.
 9. I'm not able to do some things as well as most people.
 3. I take a positive attitude towards myself.
 6. I feel I have a number of good qualities.
 2. I feel I have many good qualities.
 1. I am a useful person to have around.
 8. I feel I am a person of worth, at least on a par with others.
13. When I do a job, I do it well.
 7. I am able to do things better than most people.
10. I am usually the one who can do things best.
11. I am always the best at everything I try.

start of the program. Table 3 shows the steps taken in the ordering process and the corresponding statistics. It shows, for example, that the average variance of ranks of endorsed items for the initial order was 15.107. Then pair 5,13 was interchanged, resulting in a reduction in variance of 5.052, and so on through 11 steps. The final pair interchanged was 2,8, giving a reduction of .0005, and no pair was interchanged at step 12. It is typical that the first few interchanges result in large changes, whereas the last few result in very small changes. The resulting order is given at the bottom of section 1 of Table 3 (i.e., 12, 9, ..., 5, 4).

The program next tests to see if a break-and-rejoin is called for. It does this by moving items one at a time to the opposite end of the scale, that is, the first step tries the order 9, 3, ..., 5, 4, 12 and the next one is 3, 6, ..., 5, 4, 12, 9. Indeed, it was found here that the order essentially moving 5 and 4 to the opposite end (5, 4, 12, 9, ..., 10, 11) gave a slight improvement in ϕ. Then pairwise interchange was tried again, and two more interchanges took place, giving the order 4, 5, ..., 10, 11, as listed in Table 3. Break-and-rejoin was tried again, but no further steps were called for.

The pairwise interchange with break-and-rejoin procedures was applied to each of the scales, using random and smallest nonzero eigenvalue starting orders. Summary information is given in Table 4. The table shows the initial and final average variances and the number of interchanges required. In all six cases, one break-and-rejoin step occurred.

In all cases, the procedures resulted in an appreciable reduction in variance. Also, the eigenvector starting order led to fewer required iterations, but the reduction was not dramatic. The optimum numerical value for the eigenvector is seen to be quite far from the optimum ordinal value. This can also be seen from the initial variances. In only one case is this substantially lower than for the random order, and in one case it is slightly higher.

It can also be seen that the different starting orders resulted in different final variances in two of the three cases, reflecting different final orders as a function of the different starting orders. This is a more disturbing result because it indicates a type of "local minimum" problem.

The item orders for the best solutions for the SE and War scales are shown in Tables 2 and 5. The order seems very plausible for the SE scale and reasonable for the War scale. As a check, 17 persons familiar with psychological measurement rated the items on continua of self-esteem and attitude toward war. The Spearman rho correlation between the average rating and the position derived by the procedure from endorsement responses was .91 for SE ($\tau = .79$) and .85 for War ($\tau = .68$). Nonetheless, the fact that different solutions can occur is disturbing, although perhaps it should not be unexpected because there is no guaranteed solution to the problem.

A visual impression of the degree of fit can be obtained from Figures 1 and 2, which are the binary response matrices transposed to items × persons. In the figures, a *1* represents endorsement and a blank space represents rejection of the statement. Here, the persons have been ordered by the median rank of

Table 3
Steps in Reordering of Self-Esteem Scale

1. Interchange History

Iteration	Current Variance	Pair to be Interchanged	Improvement
1	15.0107	5,13	5.0518
2	10.0552	1,4	2.0498
3	8.0054	2,7	.5132
4	7.4922	7,9	.2224
5	7.2699	4,12	.3527
6	6.8872	1,8	.0710
7	6.8162	4,5	.0269
8	6.7892	6,8	.0165
9	6.7726	2,13	.0156
10	6.7570	1,13	.0199
11	6.7371	2,8	.0006
12	6.7365		0

Best order: 12, 9, 3, 6, 8, 2, 1, 13, 7, 10, 11, 5, 4

2. Break-and-Rejoin Tests

Beginning Item	Variance if Break
9	7.7688
3	9.9877
6	14.4751
8	17.8281
2	19.4221
1	19.0181
13	18.6449
7	12.9724
10	9.8472
11	7.5673
5*	6.5864
4	6.7384
12	6.7365

*Best place to break and rejoin.

3. Interchanges After Break-and-Rejoin

Iteration	Current Variance	Pair to be Interchanged	Improvement
1	6.5864	1,8	.0109
2	6.5756	1,2	.0040
3	6.5716	4,5	.0001
4	6.5715		0

Break-and-Rejoin test shows no further improvement.
Best Order: 4, 5, 12, 9, 3, 6, 2, 1, 8, 13, 7, 10, 11

Table 4
Summary Statistics From First Questionnaire Study

Scale and Start Order	Initial Variance	Final Variance	Total Interchanges	Break and Rejoins
War				
Random	21.83	8.31	19	1
Min. Eigen	18.91	8.83	16	1
Self-Esteem				
Random	15.12	6.57	16	1
Min. Eigen	7.19	6.57	16	1
Ambiguity				
Random	62.45	50.27	40	1
Min. Eigen	63.58	50.26	31	1

the items endorsed by each. The approximation to parallelogram form is quite good for the SE scale and reasonably good for War (recall that the former gave only a single final solution).

Archeological Data

As noted above, methods like the present one can be used for "seriation" of archeological data. That is, a series of archeological sites can be placed in presumed temporal order by comparing the types of artifacts that are found in them. The data from the "Munsingen-Rain" graves (Kendall, 1971) provide

Table 5
Ordered War Scale

14. I would immediately go to war and would do everything in my power to influence others to do the same.
 6. I would participate in war to defend a foreign country against its enemy.
 2. In time of war I would disregard any possible exemptions and enlist immediately.
 3. In time of war I would not enlist but would give whatever financial aid I could.
15. In time of war I would offer my services in whatever capacity I could.
 4. In time of war I would support my country even against my convictions.
 1. I would participate in a war only if I felt the cause was just.
 7. In time of war, I would wait a month or two if possible before I would enlist.
16. I would fight in a war only if my country were invaded by the enemy.
13. I would not go to war unless I were drafted.
11. I would immediately attempt to find some technicality on which to evade going to war.
 5. I would participate in war only if my family were being threatened.
 8. I would refuse to participate in any way in war.
 9. I would rather be called a coward than go to war.
17. I would not only refuse to participate in any way in war but also attempt to influence public opinion against war.
10. I would take part in war only to avoid social ostracism.
12. I would go to war only if my friends went to war.

Figure 1
Ordered Self-Esteem Response Matrix
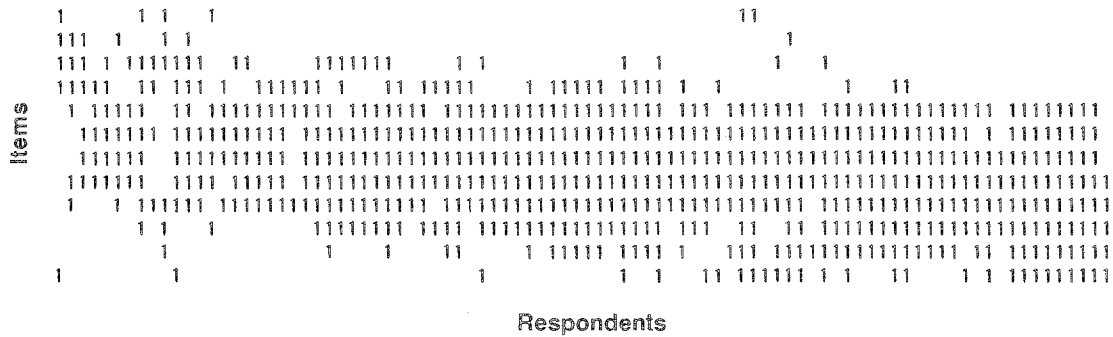
an example of this technique. For 70 varieties of artifacts (fibulae, anklets, bracelets, etc.) and 59 tombs, the response matrix shows a *1* if the artifact type was found in the tomb.

Kendall (1971) analyzed these data by means of an adaptation of multidimensional scaling of X'X. A two-dimensional analysis was performed in which the points were found to lie rather closely along a curve. The order selected by Kendall is the order of the points along a curve drawn through them. The results of this analysis were input to the present procedure, first ordering the tombs. The average initial variance was 12.19 in this rather sparse matrix. The present method reduced this variance to 8.41, a 31% reduction.

Kendall applied only an informal procedure to order the larger dimension, varieties. With this starting configuration, the average variance was 32.56; the present procedure reduced this figure to 18.82, a 42% reduction. Neither of the improvements is trivial given that the initial orders were quite good. Indeed, the multidimensional scaling order was apparently widely accepted by the literature. The ordered matrix is shown in Figure 3.

When one dimension of the matrix is quite large, it is not very feasible in terms of computer time to apply pairwise interchange. The time increases as roughly the cube of the order. Thus, some alternative

Figure 2
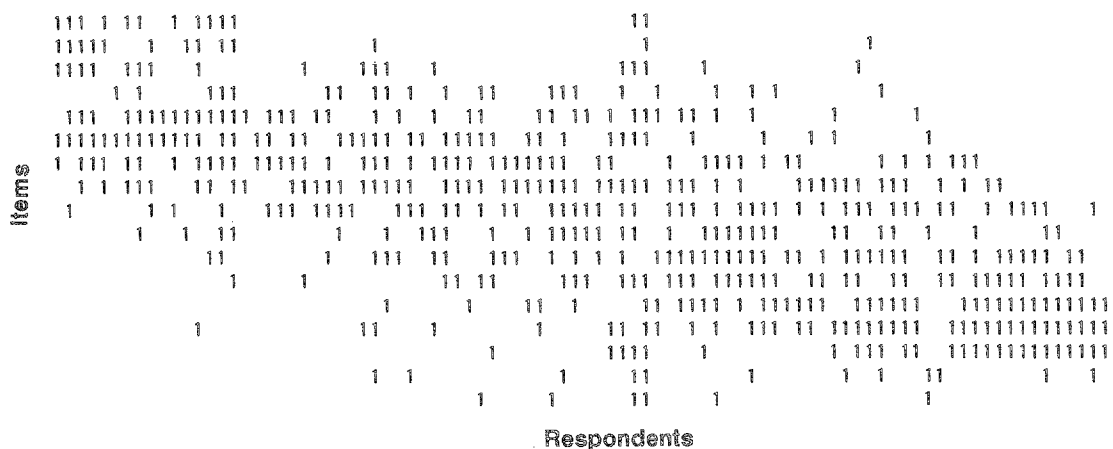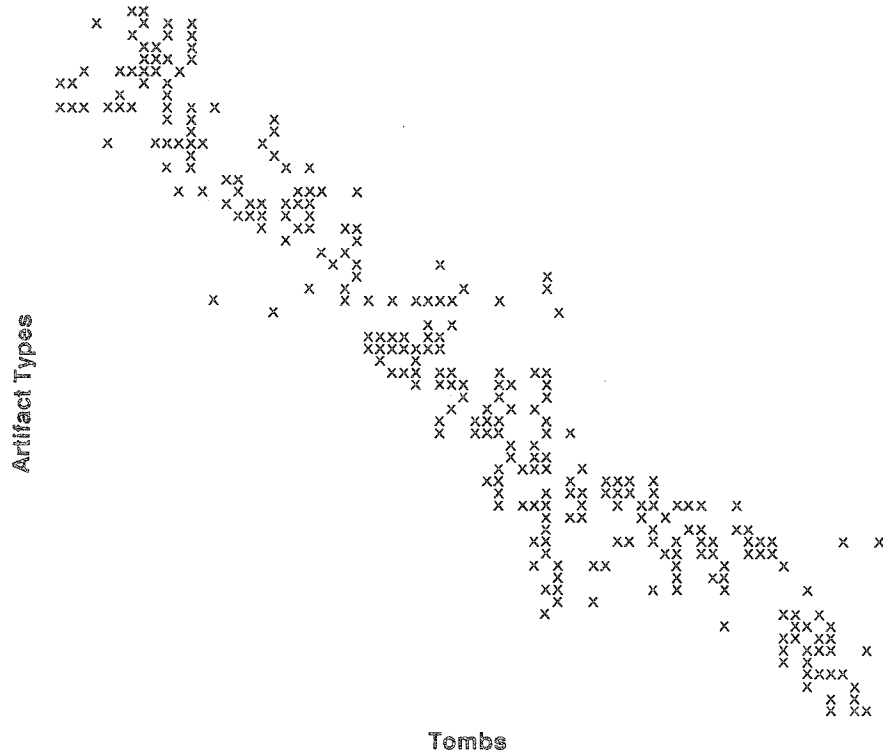Ordered Attitude Toward War Response Matrix

Figure 3
Ordered Tombs and Artifact Types

is desirable. A simple one is to apply the interchange algorithm to the smaller dimension of the matrix and then order the other dimension in terms of the mean or median rank of the endorsed items in it. This is similar in spirit to Thurstone attitude scaling (Torgerson, 1958), in which attitude items are scaled along a continuum by judges before the questionnaire is constructed. The attitudes of the respondents are then defined as the mean scale values of the endorsed items.

This method was applied to the varieties (shorter) dimension of the Munsingen-Rain data. It yielded an average variance of 22.78, not very different in this example from the minimum found using the algorithm developed here and considerably better than the initial value for the order shown by Kendall. Thus, if a good order is found for one dimension of a matrix, this seems to be a promising method for ordering the other. This idea is further discussed below.

Second Questionnaire Study

The method was adapted to include multicategory response information. Recall that the questionnaire responses were on a four-point scale representing levels of agreement, and that this was dichotomized into agree-disagree with some consequent loss of information.

A simple way to include all of the information about levels of response was developed. This involved creating a separate response matrix for each response boundary, and then reordering all of the matrices

simultaneously. For the current four-level case, this meant dichotomizing responses at Strongly Agree versus the others, at Strongly Agree or Tend to Agree versus the other two, and at Strongly Disagree versus the other three. The columns of the three matrices were interchanged simultaneously. For k levels of response, this amounts to column-ordering a matrix that is $(k - 1)(n) \times p$. Essentially, it weights the higher levels of agreement more than the lower ones. (If there are 0 responses of Strongly Agree for a person, that row of the matrix is deleted.) Correspondingly, for row-ordering the data, the matrix would be $n \times (k - 1)(p)$.

Another innovation was the use of two additional starting orders. One of these was nearest-neighbor ordering, as described above. The second was based on the observation that the final orders in the earlier study sometimes resembled the eigenvector corresponding to the second smallest nonzero eigenvalue about as well as they did the one for the smallest. Thus, the order of the entries in the eigenvector corresponding to the next-to-smallest nonzero eigenvalue was a possible starting configuration.

These modifications were implemented in the SAS version of the procedure and used in a second study. Data were gathered and analyzed using the procedures described earlier. The results will not be described here in any detail, as they are similar to those for the first study (see Gallipeau, 1984, for details). If anything, multiple solutions from different starting configurations were more frequent.

Some additional insights did emerge. For example, some cases were found where solutions differed only by a permutation of three items. In these cases it was noted that interchanging any two of the three would result in an increase in ϕ, but following this permutation with a second one permitted a decrease, leading to the solution with the smaller ϕ. Additional information was gained when each solution was plotted against the one with the smallest variance. In several cases a configuration similar to that in Figure 4 resulted. Note the cruciform nature of the plot. This means that there are two subsets of points, one of which is in the same order as the best order, and one which is in the opposite order. Apparently, it

Figure 4
Two Final Orders for the Self-Esteem Scale
Plotted Against Each Other
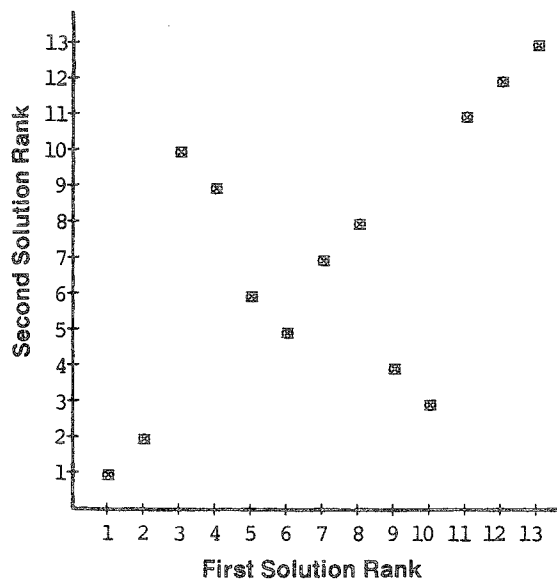
is quite common for the procedure to reach a position like this. If it does, then it may be impossible to escape from it using a pairwise interchange.

Even when the orders were rather different, the final variances were almost the same. That is, there seemed to be several solutions with nearly the same average variance but a rather different character. This means that it is important to use several starting orders and select the solution that seems to be the best compromise between low average variance and high interpretability.

Discussion

The potential applications of the method described here are considerable. It provides a possible "nonparametric" treatment of a type of data with which many investigators are faced and which they are now forced to treat by more traditional psychometric methods. "Nonparametric" appears here in quotes because the criterion function is clearly not purely nonparametric: It employs variances and it is applied to rank numbers, which comprise only one of many possible indicators of order.

The terminology (and the methodology) can be justified on several grounds. First, it is relatively nonparametric, and indeed many researchers would feel that it *is* nonparametric. More important are some pragmatic considerations. One of these is that it provided solutions that appear "better" than, for example, the more purely nonparametric nearest-neighbor analysis. The latter became relegated to the status of providing a starting configuration. Also, while considerable effort was given to inventing alternatives, as described above, none were found that provided as clear and unconfounded a description of the "bunchiness" of the data. Using variance has some computational advantages. It also seems doubtful that many of the simplifications used here would be feasible using other criteria.

The fact that final positions were sometimes variable, depending on starting orders, was somewhat disappointing, but this problem is shared with other model-fitting methods, including some where it is unrecognized or ignored. A refinement that was not implemented in the research described here is suggested by the War results shown in Figure 2. Recall that in two-set proximity data the method is employed twice, once for rows and once for columns, and note that in the last two lines of the figure the *I*s are not bunched toward the end, as the most extreme items should be in a good solution. They have been endorsed, that is, by persons who come from various parts of the continuum. (It also happened that these two items are the ones that are farthest out of their positions as ranked by the judges.)

This suggests that the mean or median rank of the persons who endorse the item could be computed and used as an additional starting order for reordering the items. Note that this tactic might be used repeatedly, and would be particularly useful when one dimension of the matrix is long (e.g., many persons). For example, if there are many more persons than items, the smaller dimension of the matrix would be ordered, and then the mean or median rank of the items endorsed by each person would be used as one starting order for persons. After one or several person-orders have been found, the average ranks of the persons endorsing each item can be used to provide new starting orders for items, and the process could continue for several iterations.

Most of the applications of the method to date have involved two-set proximity data, that is, questionnaires. However, it may be applied equally well to the other three quadrants of Coombs' (1964) system. Obviously, if the data are of the one-set proximity type and the response matrix is symmetric, then the method applies. However, it also applies to one-set proximity data that are not necessarily symmetric, such as same-different judgments where, with unidimensional stimuli, the row and column orders would ideally be the same, even though the data are not necessarily symmetric. The function would need to be altered so that simultaneous consideration could be given to variance within a given row and column. The method can also be applied to dominance data where the ideal form is the upper

triangular form that is characteristic of a Guttman scale or a simple order, because here the principle of reordering rows or columns to bunch together the positive responses will be appropriate.

One desirable addition to this method would be an index that summarizes the degree of fit to the parallelogram model. Within a given analysis, the average variance serves this purpose, but a coefficient describing the absolute degree of fit would be preferable. The present authors expended considerable effort on this topic, but were unable to find any procedure or coefficient that was sufficiently free of irrelevant influences, such as the size of the data matrix or the relative number of positive responses. This is another issue on which further research would be desirable.

References

Baker, F. B., & Hubert, L. J. (1977). Applications of combinatorial programming to data analysis: Seriation using asymmetric proximity measures. *British Journal of Mathematical and Statistical Psychology, 30,* 154–164.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology, 31,* 607–649.

Cliff, N. (1973). Psychometrics. In B. B. Wolman (Ed.), *Handbook of psychology.* New York: Prentice-Hall.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Gallipeau, D. P. (1984). *Item ordering characteristics of a parallelogram fitting procedure.* Unpublished master's thesis, University of Southern California.

Gelfand, A. E. (1971). Rapid seriation methods with archeological applications. In F. R. Hodson, D. G. Kendall, & P. Tautu (Eds.), *Mathematics in the archeological and historical sciences.* Edinburgh: Edinburgh University Press.

Goodman, L., & Kruskal, W. B. (1954). Measures of association for cross-classification. *Journal of the American Statistical Association, 49,* 732–764.

Guttman, L. (1971). Measurement as structural theory. *Psychometrika, 36,* 329–347.

Heiser, W. J. (1981). *Unfolding analysis of proximity data.* Unpublished doctoral dissertation, Leiden University, Leiden, The Netherlands.

Hubert, L. J. (1974). Problems of seriation using a subject by item response matrix. *Psychological Bulletin, 81,* 976–983.

Kendall, D. G. (1971). Seriation from abundance matrices. In F. R. Hodson, D. G. Kendall, & P. Tautu (Eds.), *Mathematics in the archeological and histor-ical sciences.* Edinburgh: Edinburgh University Press.

Levine, J. H. (1979). Joint-space analysis of "pick-any" data: Analysis of choices from an unconstrained set of alternatives. *Psychometrika, 44,* 85–92.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Robinson, J. P., & Shauer, P. R. (Eds.) (1973). *Measures of social psychological attitudes* (rev. ed.). Ann Arbor MI: Survey Research Center, Institute for Social Research.

Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika, 50,* 91–119.

Torgerson, W. (1958). *Theory and methods of scaling.* New York: Wiley.

Acknowledgments

Author's Address

Send requests for reprints or further information to Norman Cliff, Department of Psychology, University of Southern California, Los Angeles CA 90089-1061, U.S.A.